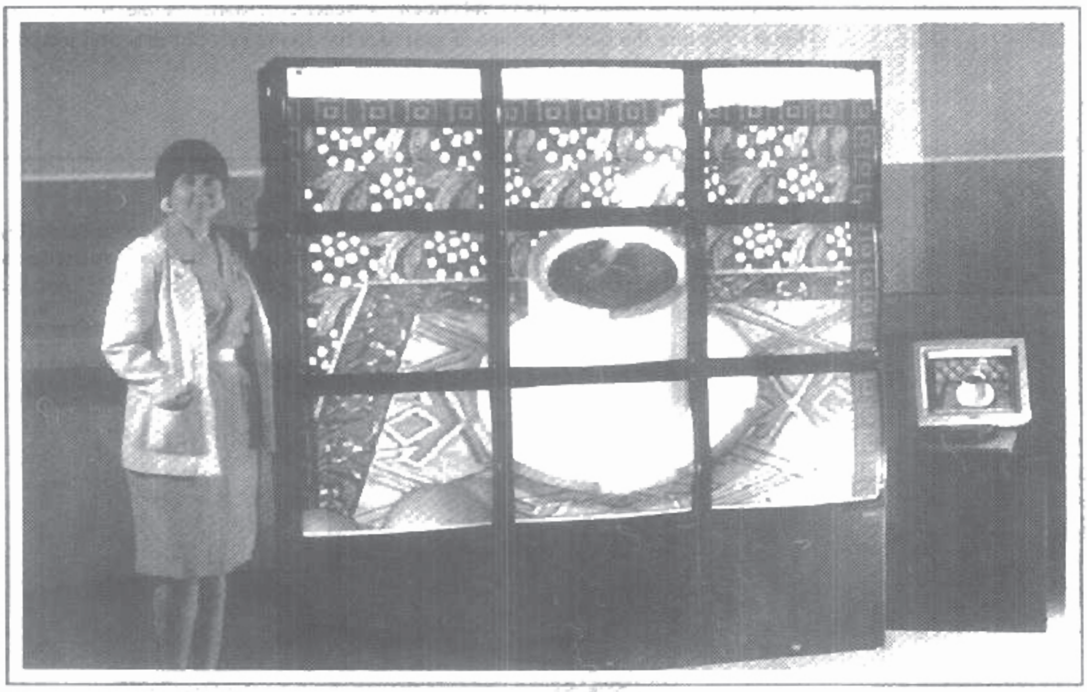
CHAPTER

2 Overview of Graphics Systems



Due to the widespread recognition of the power and utility of computer graphics in virtually all fields, a broad range of graphics hardware and software systems is now available. Graphics capabilities for both two-dimensional and three-dimensional applications are now common on general-purpose computers, including many hand-held calculators. With personal computers, we can use a wide variety of interactive input devices and graphics software packages. For higher-quality applications, we can choose from a number of sophisticated special-purpose graphics hardware systems and technologies. In this chapter, we explore the basic features of graphics hardware components and graphics software packages.

2-1 VIDEO DISPLAY DEVICES

Typically, the primary output device in a graphics system is a video monitor (Fig. 2-1). The operation of most video monitors is based on the standard **cathode-ray tube (CRT)** design, but several other technologies exist and solid-state monitors may eventually predominate.

Figure 2-1
A computer graphics workstation. *(Courtesy of Tektronix, Inc.)*

Refresh Cathode-Ray Tubes

Figure 2-2 illustrates the basic operation of a CRT. A beam of electrons (*cathode rays*), emitted by an electron gun, passes through focusing and deflection systems that direct the beam toward specified positions on the phosphor-coated screen. The phosphor then emits a small spot of light at each position contacted by the electron beam. Because the light emitted by the phosphor fades very rapidly, some method is needed for maintaining the screen picture. One way to keep the phosphor glowing is to redraw the picture repeatedly by quickly directing the electron beam back over the same points. This type of display is called a **refresh CRT**.

The primary components of an electron gun in a CRT are the heated metal cathode and a control grid (Fig. 2-3). Heat is supplied to the cathode by directing a current through a coil of wire, called the filament, inside the cylindrical cathode structure. This causes electrons to be "boiled off" the hot cathode surface. In the vacuum inside the CRT envelope, the free, negatively charged electrons are then accelerated toward the phosphor coating by a high positive voltage. The acceler-

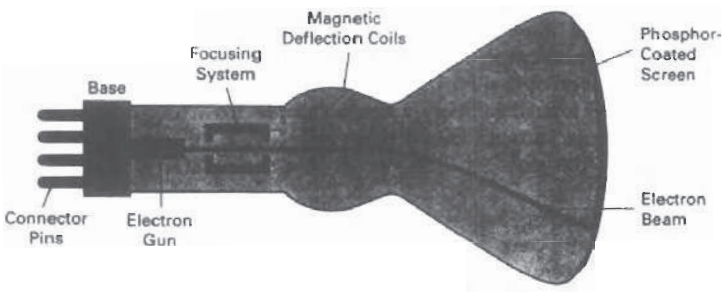


Figure 2-2
Basic design of a magnetic-deflection CRT.

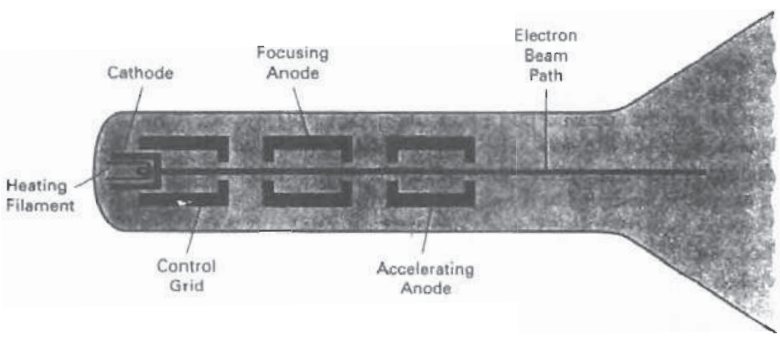


Figure 2-3
Operation of an electron gun with an accelerating anode.

ating voltage can be generated with a positively charged metal coating on the inside of the CRT envelope near the phosphor screen, or an accelerating anode can be used, as in Fig. 2-3. Sometimes the electron gun is built to contain the accelerating anode and focusing system within the same unit.

Intensity of the electron beam is controlled by setting voltage levels on the control grid, which is a metal cylinder that fits over the cathode. A high negative voltage applied to the control grid will shut off the beam by repelling electrons and stopping them from passing through the small hole at the end of the control grid structure. A smaller negative voltage on the control grid simply decreases the number of electrons passing through. Since the amount of light emitted by the phosphor coating depends on the number of electrons striking the screen, we control the brightness of a display by varying the voltage on the control grid. We specify the intensity level for individual screen positions with graphics software commands, as discussed in Chapter 3.

The focusing system in a CRT is needed to force the electron beam to converge into a small spot as it strikes the phosphor. Otherwise, the electrons would repel each other, and the beam would spread out as it approaches the screen. Focusing is accomplished with either electric or magnetic fields. Electrostatic focusing is commonly used in television and computer graphics monitors. With electrostatic focusing, the electron beam passes through a positively charged metal cylinder that forms an electrostatic lens, as shown in Fig. 2-3. The action of the electrostatic lens focuses the electron beam at the center of the screen, in exactly the same way that an optical lens focuses a beam of light at a particular focal distance. Similar lens focusing effects can be accomplished with a magnetic field set up by a coil mounted around the outside of the CRT envelope. Magnetic lens focusing produces the smallest spot size on the screen and is used in special-purpose devices.

Additional focusing hardware is used in high-precision systems to keep the beam in focus at all screen positions. The distance that the electron beam must travel to different points on the screen varies because the radius of curvature for most CRTs is greater than the distance from the focusing system to the screen center. Therefore, the electron beam will be focused properly only at the center of the screen. As the beam moves to the outer edges of the screen, displayed images become blurred. To compensate for this, the system can adjust the focusing according to the screen position of the beam.

As with focusing, deflection of the electron beam can be controlled either with electric fields or with magnetic fields. Cathode-ray tubes are now commonly constructed with magnetic deflection coils mounted on the outside of the CRT envelope, as illustrated in Fig. 2-2. Two pairs of coils are used, with the coils in each pair mounted on opposite sides of the neck of the CRT envelope. One pair is mounted on the top and bottom of the neck, and the other pair is mounted on opposite sides of the neck. The magnetic field produced by each pair of coils results in a transverse deflection force that is perpendicular both to the direction of the magnetic field and to the direction of travel of the electron beam. Horizontal deflection is accomplished with one pair of coils, and vertical deflection by the other pair. The proper deflection amounts are attained by adjusting the current through the coils. When electrostatic deflection is used, two pairs of parallel plates are mounted inside the CRT envelope. One pair of plates is mounted horizontally to control the vertical deflection, and the other pair is mounted vertically to control horizontal deflection (Fig. 2-4).

Spots of light are produced on the screen by the transfer of the CRT beam energy to the phosphor. When the electrons in the beam collide with the phos-

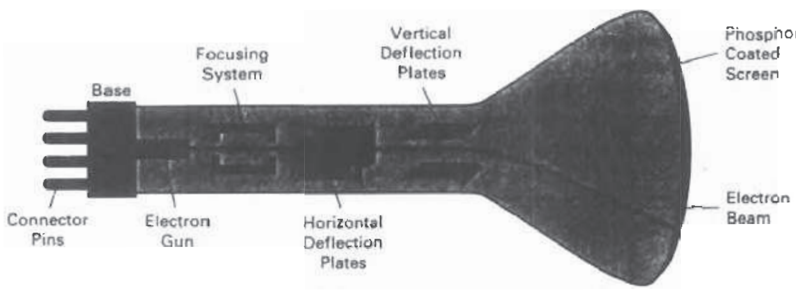


Figure 2-4
Electrostatic deflection of the electron beam in a CRT.

phor coating, they are stopped and their kinetic energy is absorbed by the phosphor. Part of the beam energy is converted by friction into heat energy, and the remainder causes electrons in the phosphor atoms to move up to higher quantum-energy levels. After a short time, the "excited" phosphor electrons begin dropping back to their stable ground state, giving up their extra energy as small quantums of light energy. What we see on the screen is the combined effect of all the electron light emissions: a glowing spot that quickly fades after all the excited phosphor electrons have returned to their ground energy level. The frequency (or color) of the light emitted by the phosphor is proportional to the energy difference between the excited quantum state and the ground state.

Different kinds of phosphors are available for use in a CRT. Besides color, a major difference between phosphors is their **persistence**: how long they continue to emit light (that is, have excited electrons returning to the ground state) after the CRT beam is removed. Persistence is defined as the time it takes the emitted light from the screen to decay to one-tenth of its original intensity. Lower-persistence phosphors require higher refresh rates to maintain a picture on the screen without flicker. A phosphor with low persistence is useful for animation; a high-persistence phosphor is useful for displaying highly complex, static pictures. Although some phosphors have a persistence greater than 1 second, graphics monitors are usually constructed with a persistence in the range from 10 to 60 microseconds.

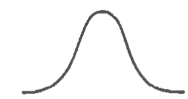


Figure 2-5
Intensity distribution of an illuminated phosphor spot on a CRT screen.

Figure 2-5 shows the intensity distribution of a spot on the screen. The intensity is greatest at the center of the spot, and decreases with a Gaussian distribution out to the edges of the spot. This distribution corresponds to the cross-sectional electron density distribution of the CRT beam.

The maximum number of points that can be displayed without overlap on a CRT is referred to as the **resolution**. A more precise definition of resolution is the number of points per centimeter that can be plotted horizontally and vertically, although it is often simply stated as the total number of points in each direction. Spot intensity has a Gaussian distribution (Fig. 2-5), so two adjacent spots will appear distinct as long as their separation is greater than the diameter at which each spot has an intensity of about 60 percent of that at the center of the spot. This overlap position is illustrated in Fig. 2-6. Spot size also depends on intensity. As more electrons are accelerated toward the phospher per second, the CRT beam diameter and the illuminated spot increase. In addition, the increased excitation energy tends to spread to neighboring phosphor atoms not directly in the

Figure 2-6
Two illuminated phosphor spots are distinguishable when their separation is greater than the diameter at which a spot intensity has fallen to 60 percent of maximum.

path of the beam, which further increases the spot diameter. Thus, resolution of a CRT is dependent on the type of phosphor, the intensity to be displayed, and the focusing and deflection systems. Typical resolution on high-quality systems is 1280 by 1024, with higher resolutions available on many systems. High-resolution systems are often referred to as *high-definition systems.* The physical size of a graphics monitor is given as the length of the screen diagonal, with sizes varying from about 12 inches to 27 inches or more. A CRT monitor can be attached to a variety of computer systems, so the number of screen points that can actually be plotted depends on the capabilities of the system to which it is attached.

Another property of video monitors is **aspect ratio**. This number gives the ratio of vertical points to horizontal points necessary to produce equal-length lines in both directions on the screen. (Sometimes aspect ratio is stated in terms of the ratio of horizontal to vertical points.) An aspect ratio of 3/4 means that a vertical line plotted with three points has the same length as a horizontal line plotted with four points.

Raster-Scan Displays

The most common type of graphics monitor employing a CRT is the **raster-scan** display, based on television technology. In a raster-scan system, the electron beam is swept across the screen, one row at a time from top to bottom. As the electron beam moves across each row, the beam intensity is turned on and off to create a pattern of illuminated spots. Picture definition is stored in a memory area called the **refresh buffer** or **frame buffer**. This memory area holds the set of intensity values for all the screen points. Stored intensity values are then retrieved from the refresh buffer and "painted" on the screen one row (**scan line**) at a time (Fig. 2-7). Each screen point is referred to as a **pixel** or **pel** (shortened forms of **picture element**). The capability of a raster-scan system to store intensity information for each screen point makes it well suited for the realistic display of scenes containing subtle shading and color patterns. Home television sets and printers are examples of other systems using raster-scan methods.

Intensity range for pixel positions depends on the capability of the raster system. In a simple black-and-white system, each screen point is either on or off, so only one bit per pixel is needed to control the intensity of screen positions. For a bilevel system, a bit value of 1 indicates that the electron beam is to be turned on at that position, and a value of 0 indicates that the beam intensity is to be off. Additional bits are needed when color and intensity variations can be displayed. Up to 24 bits per pixel are included in high-quality systems, which can require several megabytes of storage for the frame buffer, depending on the resolution of the system. A system with 24 bits per pixel and a screen resolution of 1024 by 1024 requires 3 megabytes of storage for the frame buffer. On a black-and-white system with one bit per pixel, the frame buffer is commonly called a **bitmap**. For systems with multiple bits per pixel, the frame buffer is often referred to as a **pixmap**.

Refreshing on raster-scan displays is carried out at the rate of 60 to 80 frames per second, although some systems are designed for higher refresh rates. Sometimes, refresh rates are described in units of cycles per second, or Hertz (Hz), where a cycle corresponds to one frame. Using these units, we would describe a refresh rate of 60 frames per second as simply 60 Hz. At the end of each scan line, the electron beam returns to the left side of the screen to begin displaying the next scan line. The return to the left of the screen, after refreshing each

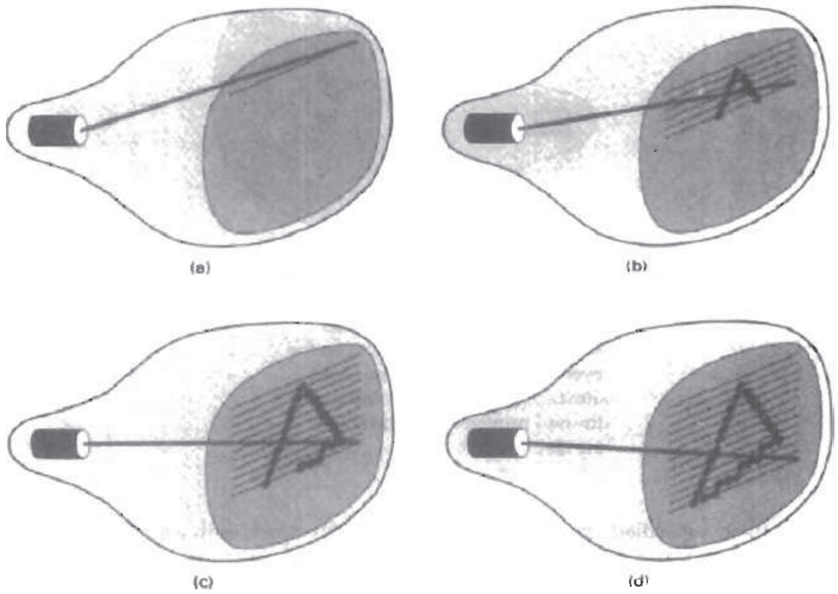


Figure 2-7
A raster-scan system displays an object as a set of discrete points across each scan line.

scan line, is called the **horizontal retrace** of the electron beam. And at the end of each frame (displayed in 1/80th to 1/60th of a second), the electron beam returns (**vertical retrace**) to the top left corner of the screen to begin the next frame.

On some raster-scan systems (and in TV sets), each frame is displayed in two passes using an *interlaced* refresh procedure. In the first pass, the beam sweeps across every other scan line from top to bottom. Then after the vertical retrace, the beam sweeps out the remaining scan lines (Fig. 2-8). Interlacing of the scan lines in this way allows us to see the entire screen displayed in one-half the time it would have taken to sweep across all the lines at once from top to bottom. Interlacing is primarily used with slower refreshing rates. On an older, 30 frame-per-second, noninterlaced display, for instance, some flicker is noticeable. But with interlacing, each of the two passes can be accomplished in 1/60th of a second, which brings the refresh rate nearer to 60 frames per second. This is an effective technique for avoiding flicker, providing that adjacent scan lines contain similar display information.

Random-Scan Displays

When operated as a **random-scan** display unit, a CRT has the electron beam directed only to the parts of the screen where a picture is to be drawn. Random-scan monitors draw a picture one line at a time and for this reason are also referred to as **vector** displays (or **stroke-writing** or **calligraphic** displays). The component lines of a picture can be drawn and refreshed by a random-scan sys-

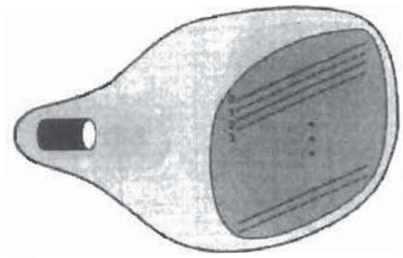


Figure 2-8
Interlacing scan lines on a raster-scan display. First, all points on the even-numbered (solid) scan lines are displayed; then all points along the odd-numbered (dashed) lines are displayed.

tem in any specified order (Fig. 2-9). A pen plotter operates in a similar way and is an example of a random-scan, hard-copy device.

Refresh rate on a random-scan system depends on the number of lines to be displayed. Picture definition is now stored as a set of line-drawing commands in an area of memory referred to as the **refresh display file**. Sometimes the refresh display file is called the **display list, display program**, or simply the **refresh buffer**. To display a specified picture, the system cycles through the set of commands in the display file, drawing each component line in turn. After all line-drawing commands have been processed, the system cycles back to the first line command in the list. Random-scan displays are designed to draw all the component lines of a picture 30 to 60 times each second. High-quality vector systems are capable of handling approximately 100,000 "short" lines at this refresh rate. When a small set of lines is to be displayed, each refresh cycle is delayed to avoid refresh rates greater than 60 frames per second. Otherwise, faster refreshing of the set of lines could burn out the phosphor.

Random-scan systems are designed for line-drawing applications and cannot display realistic shaded scenes. Since picture definition is stored as a set of line-drawing instructions and not as a set of intensity values for all screen points, vector displays generally have higher resolution than raster systems. Also, vector displays produce smooth line drawings because the CRT beam directly follows the line path. A raster system, in contrast, produces jagged lines that are plotted as discrete point sets.

Color CRT Monitors

A CRT monitor displays color pictures by using a combination of phosphors that emit different-colored light. By combining the emitted light from the different phosphors, a range of colors can be generated. The two basic techniques for producing color displays with a CRT are the beam-penetration method and the shadow-mask method.

The **beam-penetration** method for displaying color pictures has been used with random-scan monitors. Two layers of phosphor, usually red and green, are

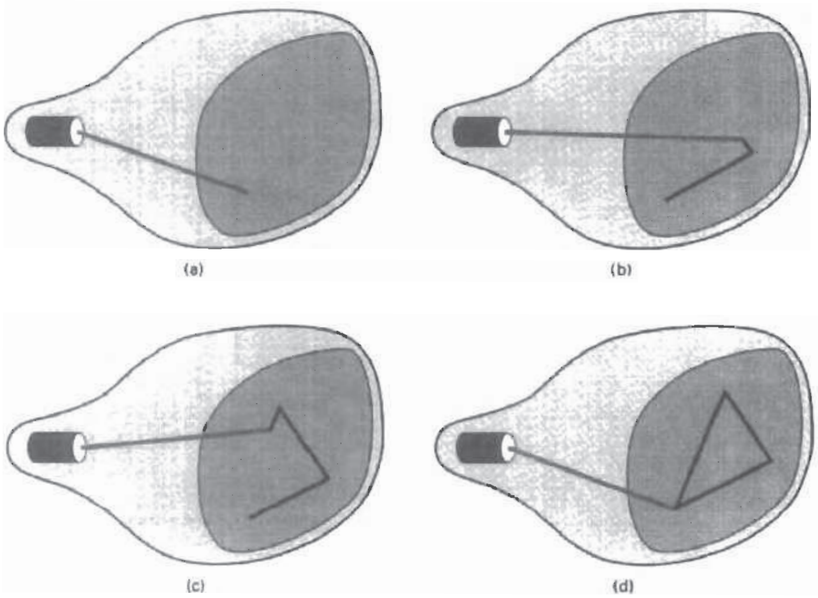


Figure 2-9
A random-scan system draws the component lines of an object in any order specified.

coated onto the inside of the CRT screen, and the displayed color depends on how far the electron beam penetrates into the phosphor layers. A beam of slow electrons excites only the outer red layer. A beam of very fast electrons penetrates through the red layer and excites the inner green layer. At intermediate beam speeds, combinations of red and green light are emitted to show two additional colors, orange and yellow. The speed of the electrons, and hence the screen color at any point, is controlled by the beam-acceleration voltage. Beam penetration has been an inexpensive way to produce color in random-scan monitors, but only four colors are possible, and the quality of pictures is not as good as with other methods.

Shadow-mask methods are commonly used in raster-scan systems (including color TV) because they produce a much wider range of colors than the beam-penetration method. A shadow-mask CRT has three phosphor color dots at each pixel position. One phosphor dot emits a red light, another emits a green light, and the third emits a blue light. This type of CRT has three electron guns, one for each color dot, and a shadow-mask grid just behind the phosphor-coated screen. Figure 2-10 illustrates the *delta-delta* shadow-mask method, commonly used in color CRT systems. The three electron beams are deflected and focused as a group onto the shadow mask, which contains a series of holes aligned with the phosphor-dot patterns. When the three beams pass through a hole in the shadow mask, they activate a dot triangle, which appears as a small color spot on the screen. The phosphor dots in the triangles are arranged so that each electron beam can activate only its corresponding color dot when it passes through the

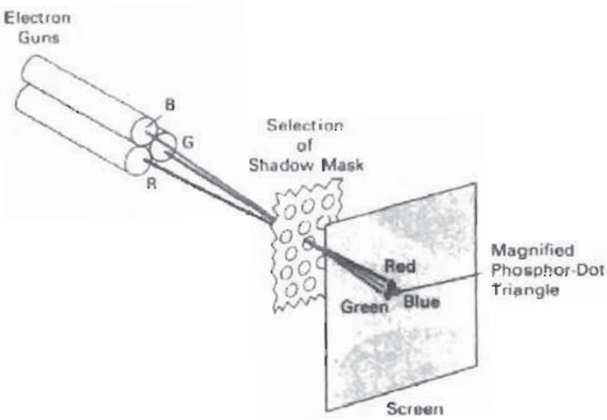


Figure 2-10
Operation of a delta-delta, shadow-mask CRT. Three electron guns, aligned with the triangular color-dot patterns on the screen, are directed to each dot triangle by a shadow mask.

shadow mask. Another configuration for the three electron guns is an *in-line* arrangement in which the three electron guns, and the corresponding red–green–blue color dots on the screen, are aligned along one scan line instead of in a triangular pattern. This in-line arrangement of electron guns is easier to keep in alignment and is commonly used in high-resolution color CRTs.

We obtain color variations in a shadow-mask CRT by varying the intensity levels of the three electron beams. By turning off the red and green guns, we get only the color coming from the blue phosphor. Other combinations of beam intensities produce a small light spot for each pixel position, since our eyes tend to merge the three colors into one composite. The color we see depends on the amount of excitation of the red, green, and blue phosphors. A white (or gray) area is the result of activating all three dots with equal intensity. Yellow is produced with the green and red dots only, magenta is produced with the blue and red dots, and cyan shows up when blue and green are activated equally. In some low-cost systems, the electron beam can only be set to on or off, limiting displays to eight colors. More sophisticated systems can set intermediate intensity levels for the electron beams, allowing several million different colors to be generated.

Color graphics systems can be designed to be used with several types of CRT display devices. Some inexpensive home-computer systems and video games are designed for use with a color TV set and an RF (radio-frequency) modulator. The purpose of the RF modulator is to simulate the signal from a broadcast TV station. This means that the color and intensity information of the picture must be combined and superimposed on the broadcast-frequency carrier signal that the TV needs to have as input. Then the circuitry in the TV takes this signal from the RF modulator, extracts the picture information, and paints it on the screen. As we might expect, this extra handling of the picture information by the RF modulator and TV circuitry decreases the quality of displayed images.

Composite monitors are adaptations of TV sets that allow bypass of the broadcast circuitry. These display devices still require that the picture informa-

tion be combined, but no carrier signal is needed. Picture information is combined into a composite signal and then separated by the monitor, so the resulting picture quality is still not the best attainable.

Color CRTs in graphics systems are designed as **RGB monitors**. These monitors use shadow-mask methods and take the intensity level for each electron gun (red, green, and blue) directly from the computer system without any intermediate processing. High-quality raster-graphics systems have 24 bits per pixel in the frame buffer, allowing 256 voltage settings for each electron gun and nearly 17 million color choices for each pixel. An RGB color system with 24 bits of storage per pixel is generally referred to as a **full-color system** or a **true-color system**.

Direct-View Storage Tubes

An alternative method for maintaining a screen image is to store the picture information inside the CRT instead of refreshing the screen. A **direct-view storage tube (DVST)** stores the picture information as a charge distribution just behind the phosphor-coated screen. Two electron guns are used in a DVST. One, the primary gun, is used to store the picture pattern; the second, the flood gun, maintains the picture display.

A DVST monitor has both disadvantages and advantages compared to the refresh CRT. Because no refreshing is needed, very complex pictures can be displayed at very high resolutions without flicker. Disadvantages of DVST systems are that they ordinarily do not display color and that selected parts of a picture cannot be erased. To eliminate a picture section, the entire screen must be erased and the modified picture redrawn. The erasing and redrawing process can take several seconds for a complex picture. For these reasons, storage displays have been largely replaced by raster systems.

Flat-Panel Displays

Although most graphics monitors are still constructed with CRTs, other technologies are emerging that may soon replace CRT monitors. The term **flat-panel display** refers to a class of video devices that have reduced volume, weight, and power requirements compared to a CRT. A significant feature of flat-panel displays is that they are thinner than CRTs, and we can hang them on walls or wear them on our wrists. Since we can even write on some flat-panel displays, they will soon be available as pocket notepads. Current uses for flat-panel displays include small TV monitors, calculators, pocket video games, laptop computers, armrest viewing of movies on airlines, as advertisement boards in elevators, and as graphics displays in applications requiring rugged, portable monitors.

We can separate flat-panel displays into two categories: **emissive displays** and **nonemissive displays**. The emissive displays (or **emitters**) are devices that convert electrical energy into light. Plasma panels, thin-film electroluminescent displays, and light-emitting diodes are examples of emissive displays. Flat CRTs have also been devised, in which electron beams are accelerated parallel to the screen, then deflected 90° to the screen. But flat CRTs have not proved to be as successful as other emissive devices. Nonemmissive displays (or **nonemitters**) use optical effects to convert sunlight or light from some other source into graphics patterns. The most important example of a nonemissive flat-panel display is a liquid-crystal device.

Plasma panels, also called **gas-discharge displays**, are constructed by filling the region between two glass plates with a mixture of gases that usually in-

cludes neon. A series of vertical conducting ribbons is placed on one glass panel, and a set of horizontal ribbons is built into the other glass panel (Fig. 2-11). Firing voltages applied to a pair of horizontal and vertical conductors cause the gas at the intersection of the two conductors to break down into a glowing plasma of electrons and ions. Picture definition is stored in a refresh buffer, and the firing voltages are applied to refresh the pixel positions (at the intersections of the conductors) 60 times per second. Alternating-current methods are used to provide faster application of the firing voltages, and thus brighter displays. Separation between pixels is provided by the electric field of the conductors. Figure 2-12 shows a high-definition plasma panel. One disadvantage of plasma panels has been that they were strictly monochromatic devices, but systems have been developed that are now capable of displaying color and grayscale.

Thin-film electroluminescent displays are similar in construction to a plasma panel. The difference is that the region between the glass plates is filled with a phosphor, such as zinc sulfide doped with manganese, instead of a gas (Fig. 2-13). When a sufficiently high voltage is applied to a pair of crossing electrodes, the phosphor becomes a conductor in the area of the intersection of the two electrodes. Electrical energy is then absorbed by the manganese atoms, which then release the energy as a spot of light similar to the glowing plasma effect in a plasma panel. Electroluminescent displays require more power than plasma panels, and good color and gray scale displays are hard to achieve.

A third type of emissive device is the **light-emitting diode (LED)**. A matrix of diodes is arranged to form the pixel positions in the display, and picture definition is stored in a refresh buffer. As in scan-line refreshing of a CRT, information

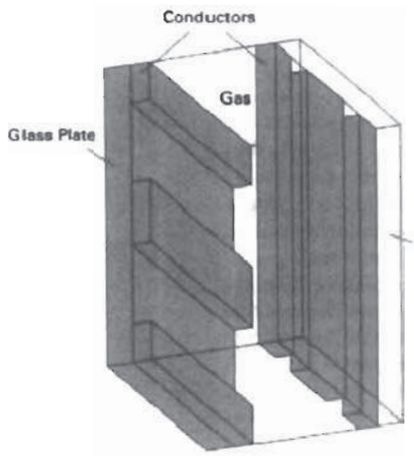


Figure 2-11
Basic design of a plasma-panel display device.

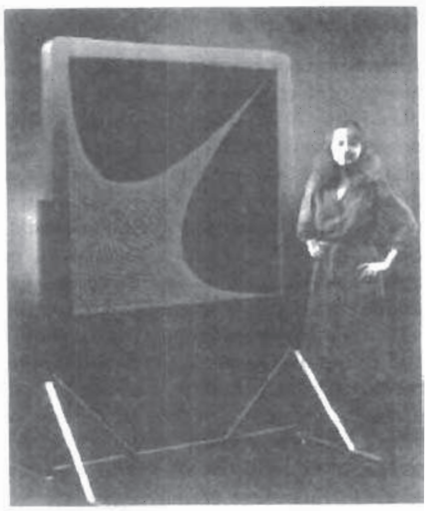


Figure 2-12
A plasma-panel display with a resolution of 2048 by 2048 and a screen diagonal of 1.5 meters.
(*Courtesy of Photonics Systems.*)

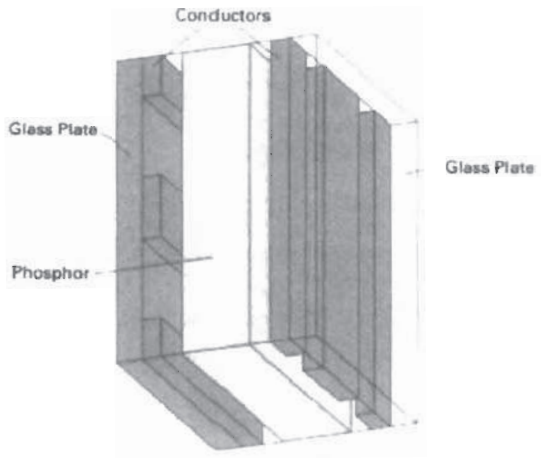


Figure 2-13
Basic design of a thin-film electroluminescent display device.

is read from the refresh buffer and converted to voltage levels that are applied to the diodes to produce the light patterns in the display.

Liquid-crystal displays (LCDs) are commonly used in small systems, such as calculators (Fig. 2-14) and portable, laptop computers (Fig. 2-15). These nonemissive devices produce a picture by passing polarized light from the surroundings or from an internal light source through a liquid-crystal material that can be aligned to either block or transmit the light.

The term *liquid crystal* refers to the fact that these compounds have a crystalline arrangement of molecules, yet they flow like a liquid. Flat-panel displays commonly use nematic (threadlike) liquid-crystal compounds that tend to keep the long axes of the rod-shaped molecules aligned. A flat-panel display can then be constructed with a nematic liquid crystal, as demonstrated in Fig. 2-16. Two glass plates, each containing a light polarizer at right angles to the other plate, sandwich the liquid-crystal material. Rows of horizontal transparent conductors are built into one glass plate, and columns of vertical conductors are put into the other plate. The intersection of two conductors defines a pixel position. Normally, the molecules are aligned as shown in the "on state" of Fig. 2-16. Polarized light passing through the material is twisted so that it will pass through the opposite polarizer. The light is then reflected back to the viewer. To turn off the pixel, we apply a voltage to the two intersecting conductors to align the molecules so that the light is not twisted. This type of flat-panel device is referred to as a **passive-matrix** LCD. Picture definitions are stored in a refresh buffer, and the screen is refreshed at the rate of 60 frames per second, as in the emissive devices. Back lighting is also commonly applied using solid-state electronic devices, so that the system is not completely dependent on outside light sources. Colors can be displayed by using different materials or dyes and by placing a triad of color pixels at each screen location. Another method for constructing LCDs is to place a transistor at each pixel location, using thin-film transistor technology. The transistors are used to control the voltage at pixel locations and to prevent charge from gradually leaking out of the liquid-crystal cells. These devices are called **active-matrix** displays.

Figure 2-14
A hand calculator with an LCD screen. (*Courtesy of Texas Instruments.*)

Figure 2-15
A backlit, passive-matrix, liquid-crystal display in a laptop computer, featuring 256 colors, a screen resolution of 640 by 400, and a screen diagonal of 9 inches.
(*Courtesy of Apple Computer, Inc.*)

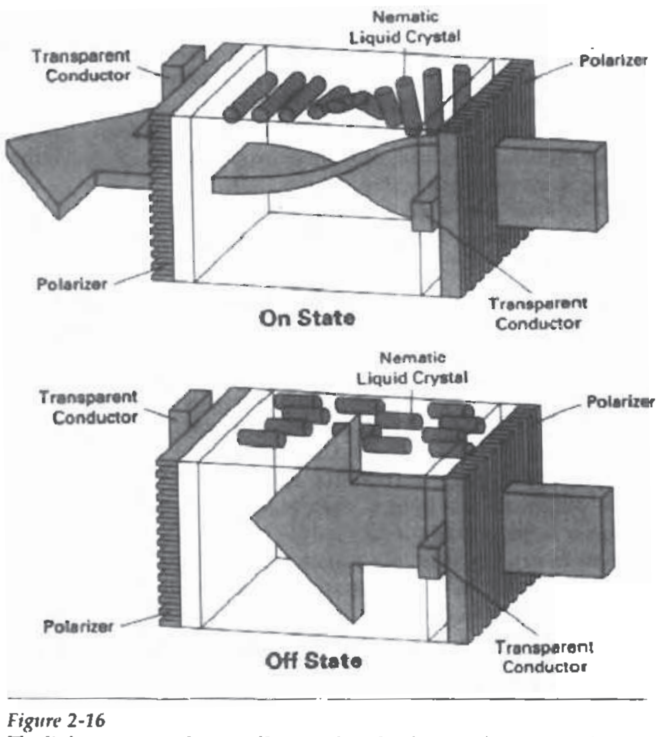


Figure 2-16
The light-twisting, shutter effect used in the design of most liquid-crystal display devices.

Three-Dimensional Viewing Devices

Graphics monitors for the display of three-dimensional scenes have been devised using a technique that reflects a CRT image from a vibrating, flexible mirror. The operation of such a system is demonstrated in Fig. 2-17. As the varifocal mirror vibrates, it changes focal length. These vibrations are synchronized with the display of an object on a CRT so that each point on the object is reflected from the mirror into a spatial position corresponding to the distance of that point from a specified viewing position. This allows us to walk around an object or scene and view it from different sides.

Figure 2-18 shows the Genisco SpaceGraph system, which uses a vibrating mirror to project three-dimensional objects into a 25-cm by 25-cm by 25-cm volume. This system is also capable of displaying two-dimensional cross-sectional "slices" of objects selected at different depths. Such systems have been used in medical applications to analyze data from ultrasonography and CAT scan devices, in geological applications to analyze topological and seismic data, in design applications involving solid objects, and in three-dimensional simulations of systems, such as molecules and terrain.

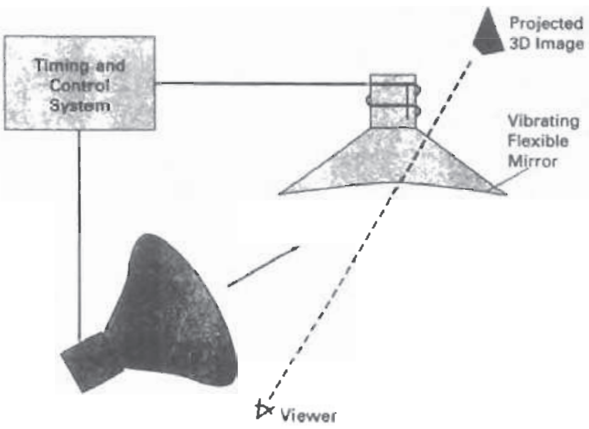


Figure 2-17
Operation of a three-dimensional display system using a vibrating mirror that changes focal length to match the depth of points in a scene.

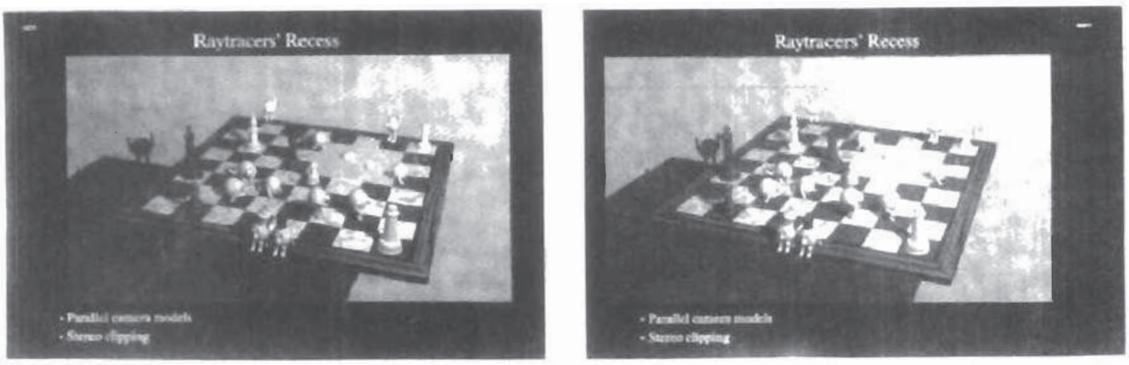
Figure 2-18
The SpaceGraph interactive graphics system displays objects in three dimensions using a vibrating, flexible mirror. *(Courtesy of Genisco Computers Corporation.)*

Stereoscopic and Virtual-Reality Systems

Another technique for representing three-dimensional objects is displaying stereoscopic views. This method does not produce true three-dimensional images, but it does provide a three-dimensional effect by presenting a different view to each eye of an observer so that scenes do appear to have depth (Fig. 2-19).

To obtain a stereoscopic projection, we first need to obtain two views of a scene generated from a viewing direction corresponding to each eye (left and right). We can construct the two views as computer-generated scenes with different viewing positions, or we can use a stereo camera pair to photograph some object or scene. When we simultaneous look at the left view with the left eye and the right view with the right eye, the two views merge into a single image and we perceive a scene with depth. Figure 2-20 shows two views of a computer-generated scene for stereographic projection. To increase viewing comfort, the areas at the left and right edges of this scene that are visible to only one eye have been eliminated.

Figure 2-19
Viewing a stereoscopic projection.
(*Courtesy of StereoGraphics Corporation.*)



Left Right

Figure 2-20
A stereoscopic viewing pair. (*Courtesy of Jerry Farm.*)

One way to produce a stereoscopic effect is to display each of the two views with a raster system on alternate refresh cycles. The screen is viewed through glasses, with each lens designed to act as a rapidly alternating shutter that is synchronized to block out one of the views. Figure 2-21 shows a pair of stereoscopic glasses constructed with liquid-crystal shutters and an infrared emitter that synchronizes the glasses with the views on the screen.

Stereoscopic viewing is also a component in **virtual-reality** systems, where users can step into a scene and interact with the environment. A headset (Fig. 2-22) containing an optical system to generate the stereoscopic views is commonly used in conjuction with interactive input devices to locate and manipulate objects in the scene. A sensing system in the headset keeps track of the viewer's position, so that the front and back of objects can be seen as the viewer

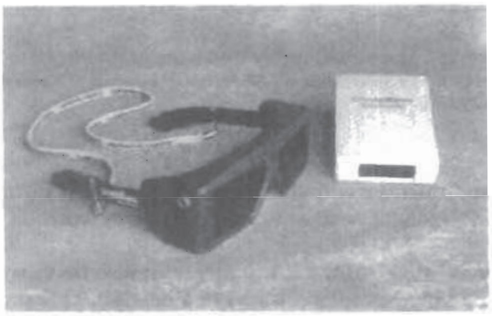


Figure 2-21
Glasses for viewing a stereoscopic scene and an infrared synchronizing emitter. (*Courtesy of StereoGraphics Corporation.*)

Figure 2-22
A headset used in virtual-reality systems. (*Courtesy of Virtual Research.*)

Figure 2-23
Interacting with a virtual-reality environment. (*Courtesy of the National Center for Supercomputing Applications, University of Illinois at Urbana-Champaign.*)

"walks through" and interacts with the display. Figure 2-23 illustrates interaction with a virtual scene, using a headset and a data glove worn on the right hand (Section 2-5).

An interactive virtual-reality environment can also be viewed with stereoscopic glasses and a video monitor, instead of a headset. This provides a means for obtaining a lower-cost virtual-reality system. As an example, Fig. 2-24 shows an ultrasound tracking device with six degrees of freedom. The tracking device is placed on top of the video display and is used to monitor head movements so that the viewing position for a scene can be changed as head position changes.

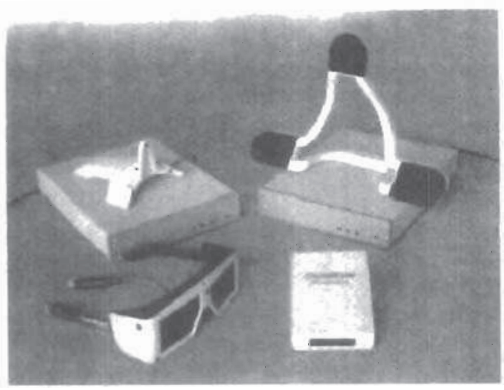


Figure 2-24
An ultrasound tracking device used with stereoscopic glasses to track head position. (*Courtesy of StereoGraphics Corporation.*)

2-2 RASTER-SCAN SYSTEMS

Interactive raster graphics systems typically employ several processing units. In addition to the central processing unit, or CPU, a special-purpose processor, called the **video controller** or **display controller**, is used to control the operation of the display device. Organization of a simple raster system is shown in Fig. 2-25. Here, the frame buffer can be anywhere in the system memory, and the video controller accesses the frame buffer to refresh the screen. In addition to the video controller, more sophisticated raster systems employ other processors as co-processors and accelerators to implement various graphics operations.

Video Controller

Figure 2-26 shows a commonly used organization for raster systems. A fixed area of the system memory is reserved for the frame buffer, and the video controller is given direct access to the frame-buffer memory.

Frame-buffer locations, and the corresponding screen positions, are referenced in Cartesian coordinates. For many graphics monitors, the coordinate ori-

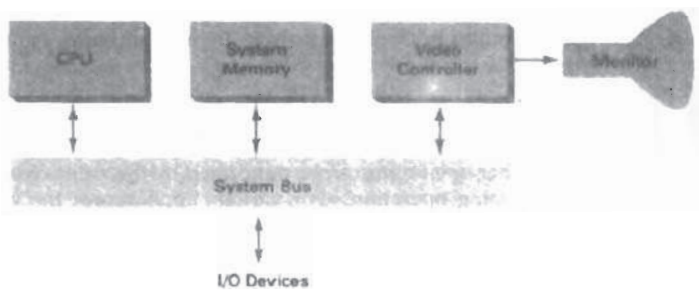


Figure 2-25
Architecture of a simple raster graphics system.

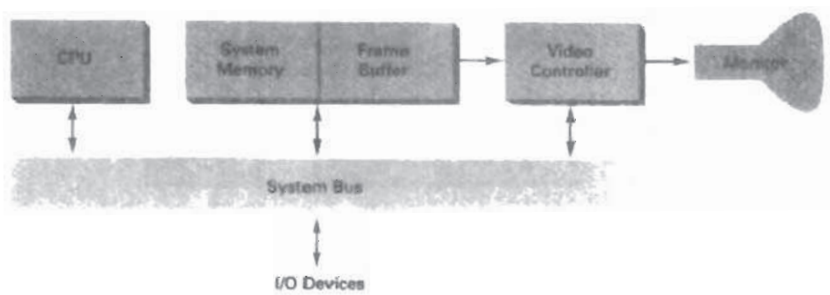


Figure 2-26
Architecture of a raster system with a fixed portion of the system memory reserved for the frame buffer.

gin is defined at the lower left screen corner (Fig. 2-27). The screen surface is then represented as the first quadrant of a two-dimensional system, with positive x values increasing to the right and positive y values increasing from bottom to top. (On some personal computers, the coordinate origin is referenced at the upper left corner of the screen, so the y values are inverted.) Scan lines are then labeled from y_{max} at the top of the screen to 0 at the bottom. Along each scan line, screen pixel positions are labeled from 0 to x_{max}.

Figure 2-27
The origin of the coordinate system for identifying screen positions is usually specified in the lower-left corner.

In Fig. 2-28, the basic refresh operations of the video controller are diagrammed. Two registers are used to store the coordinates of the screen pixels. Initially, the x register is set to 0 and the y register is set to y_{max}. The value stored in the frame buffer for this pixel position is then retrieved and used to set the intensity of the CRT beam. Then the x register is incremented by 1, and the process repeated for the next pixel on the top scan line. This procedure is repeated for each pixel along the scan line. After the last pixel on the top scan line has been processed, the x register is reset to 0 and the y register is decremented by 1. Pixels along this scan line are then processed in turn, and the procedure is repeated for each successive scan line. After cycling through all pixels along the bottom scan line ($y = 0$), the video controller resets the registers to the first pixel position on the top scan line and the refresh process starts over.

Since the screen must be refreshed at the rate of 60 frames per second, the simple procedure illustrated in Fig. 2-28 cannot be accommodated by typical RAM chips. The cycle time is too slow. To speed up pixel processing, video controllers can retrieve multiple pixel values from the refresh buffer on each pass. The multiple pixel intensities are then stored in a separate register and used to control the CRT beam intensity for a group of adjacent pixels. When that group of pixels has been processed, the next block of pixel values is retrieved from the frame buffer.

A number of other operations can be performed by the video controller, besides the basic refreshing operations. For various applications, the video con-

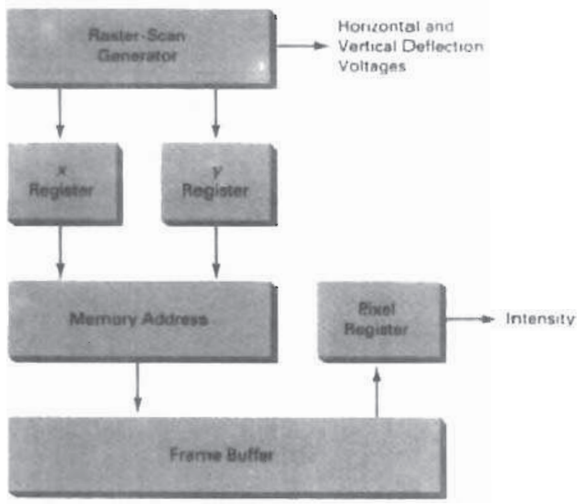


Figure 2-28
Basic video-controller refresh operations.

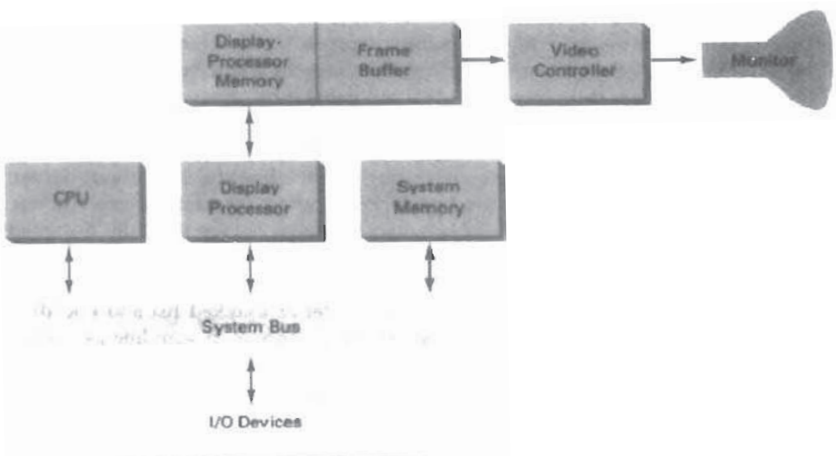


Figure 2-29
Architecture of a raster-graphics system with a display processor.

troller can retrieve pixel intensities from different memory areas on different refresh cycles. In high-quality systems, for example, two frame buffers are often provided so that one buffer can be used for refreshing while the other is being filled with intensity values. Then the two buffers can switch roles. This provides a fast mechanism for generating real-time animations, since different views of moving objects can be successively loaded into the refresh buffers. Also, some transformations can be accomplished by the video controller. Areas of the screen can be enlarged, reduced, or moved from one location to another during the refresh cycles. In addition, the video controller often contains a lookup table, so that pixel values in the frame buffer are used to access the lookup table instead of controlling the CRT beam intensity directly. This provides a fast method for changing screen intensity values, and we discuss lookup tables in more detail in Chapter 4. Finally, some systems are designed to allow the video controller to mix the frame-buffer image with an input image from a television camera or other input device.

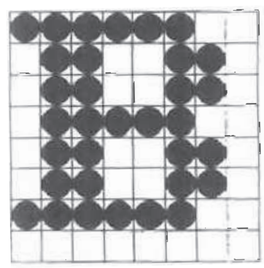


Figure 2-30
A character defined as a rectangular grid of pixel positions.

Raster-Scan Display Processor

Figure 2-29 shows one way to set up the organization of a raster system containing a separate **display processor,** sometimes referred to as a **graphics controller** or a **display coprocessor**. The purpose of the display processor is to free the CPU from the graphics chores. In addition to the system memory, a separate display-processor memory area can also be provided.

A major task of the display processor is digitizing a picture definition given in an application program into a set of pixel-intensity values for storage in the frame buffer. This digitization process is called **scan conversion**. Graphics commands specifying straight lines and other geometric objects are scan converted into a set of discrete intensity points. Scan converting a straight-line segment, for example, means that we have to locate the pixel positions closest to the line path and store the intensity for each position in the frame buffer. Similar methods are used for scan converting curved lines and polygon outlines. Characters can be defined with rectangular grids, as in Fig. 2-30, or they can be defined with curved

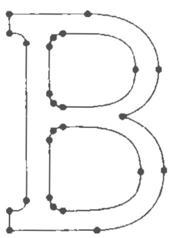


Figure 2-31
A character defined as a curve outline.

outlines, as in Fig. 2-31. The array size for character grids can vary from about 5 by 7 to 9 by 12 or more for higher-quality displays. A character grid is displayed by superimposing the rectangular grid pattern into the frame buffer at a specified coordinate position. With characters that are defined as curve outlines, character shapes are scan converted into the frame buffer.

Display processors are also designed to perform a number of additional operations. These functions include generating various line styles (dashed, dotted, or solid), displaying color areas, and performing certain transformations and manipulations on displayed objects. Also, display processors are typically designed to interface with interactive input devices, such as a mouse.

In an effort to reduce memory requirements in raster systems, methods have been devised for organizing the frame buffer as a linked list and encoding the intensity information. One way to do this is to store each scan line as a set of integer pairs. One number of each pair indicates an intensity value, and the second number specifies the number of adjacent pixels on the scan line that are to have that intensity. This technique, called **run-length encoding**, can result in a considerable saving in storage space if a picture is to be constructed mostly with long runs of a single color each. A similar approach can be taken when pixel intensities change linearly. Another approach is to encode the raster as a set of rectangular areas (**cell encoding**). The disadvantages of encoding runs are that intensity changes are difficult to make and storage requirements actually increase as the length of the runs decreases. In addition, it is difficult for the display controller to process the raster when many short runs are involved.

2-3 RANDOM-SCAN SYSTEMS

The organization of a simple random-scan (vector) system is shown in Fig. 2-32. An application program is input and stored in the system memory along with a graphics package. Graphics commands in the application program are translated by the graphics package into a display file stored in the system memory. This display file is then accessed by the display processor to refresh the screen. The display processor cycles through each command in the display file program once during every refresh cycle. Sometimes the display processor in a random-scan system is referred to as a **display processing unit** or a **graphics controller**.

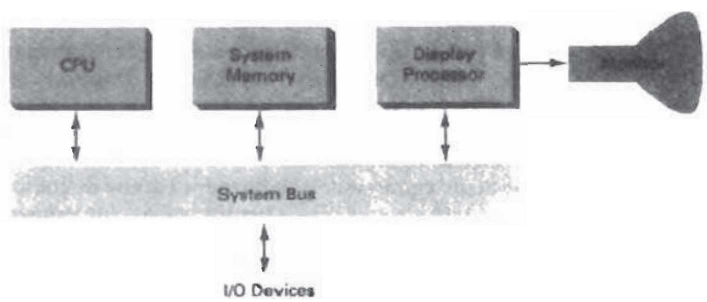


Figure 2-32
Architecture of a simple random-scan system.

Graphics patterns are drawn on a random-scan system by directing the electron beam along the component lines of the picture. Lines are defined by the values for their coordinate endpoints, and these input coordinate values are converted to x and y deflection voltages. A scene is then drawn one line at a time by positioning the beam to fill in the line between specified endpoints.

2-4 GRAPHICS MONITORS AND WORKSTATIONS

Most graphics monitors today operate as raster-scan displays, and here we survey a few of the many graphics hardware configurations available. Graphics systems range from small general-purpose computer systems with graphics capabilities (Fig. 2-33) to sophisticated full-color systems that are designed specifically for graphics applications (Fig. 2-34). A typical screen resolution for personal com-

Figure 2-33
A desktop general-purpose computer system that can be used for graphics applications. *(Courtesy of Apple Computer, Inc.)*

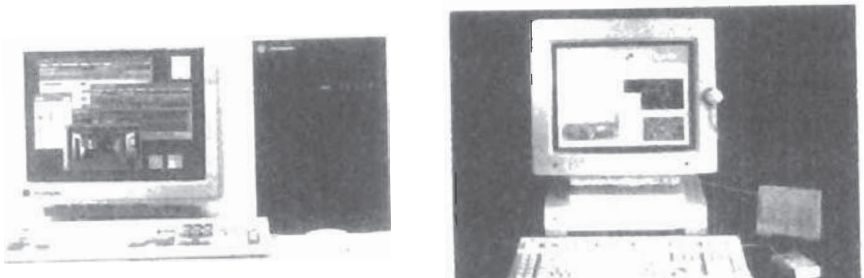


Figure 2-34
Computer graphics workstations with keyboard and mouse input devices. (a) The Iris Indigo. *(Courtesy of Silicon Graphics Corporation.)* (b) SPARCstation 10. *(Courtesy of Sun Microsystems.)*

puter systems, such as the Apple Quadra shown in Fig. 2-33, is 640 by 480, although screen resolution and other system capabilities vary depending on the size and cost of the system. Diagonal screen dimensions for general-purpose personal computer systems can range from 12 to 21 inches, and allowable color selections range from 16 to over 32,000. For workstations specifically designed for graphics applications, such as the systems shown in Fig. 2-34, typical screen resolution is 1280 by 1024, with a screen diagonal of 16 inches or more. Graphics workstations can be configured with from 8 to 24 bits per pixel (full-color systems), with higher screen resolutions, faster processors, and other options available in high-end systems.

Figure 2-35 shows a high-definition graphics monitor used in applications such as air traffic control, simulation, medical imaging, and CAD. This system has a diagonal screen size of 27 inches, resolutions ranging from 2048 by 1536 to 2560 by 2048, with refresh rates of 80 Hz or 60 Hz noninterlaced.

A multiscreen system called the MediaWall, shown in Fig. 2-36, provides a large "wall-sized" display area. This system is designed for applications that require large area displays in brightly lighted environments, such as at trade shows, conventions, retail stores, museums, or passenger terminals. MediaWall operates by splitting images into a number of sections and distributing the sections over an array of monitors or projectors using a graphics adapter and satellite control units. An array of up to 5 by 5 monitors, each with a resolution of 640 by 480, can be used in the MediaWall to provide an overall resolution of 3200 by 2400 for either static scenes or animations. Scenes can be displayed behind mullions, as in Fig. 2-36, or the mullions can be eliminated to display a continuous picture with no breaks between the various sections.

Many graphics workstations, such as some of those shown in Fig. 2-37, are configured with two monitors. One monitor can be used to show all features of an object or scene, while the second monitor displays the detail in some part of the picture. Another use for dual-monitor systems is to view a picture on one monitor and display graphics options (menus) for manipulating the picture components on the other monitor.

Figure 2-35
A very high-resolution (2560 by 2048) color monitor. (*Courtesy of BARCO Chromatics.*)

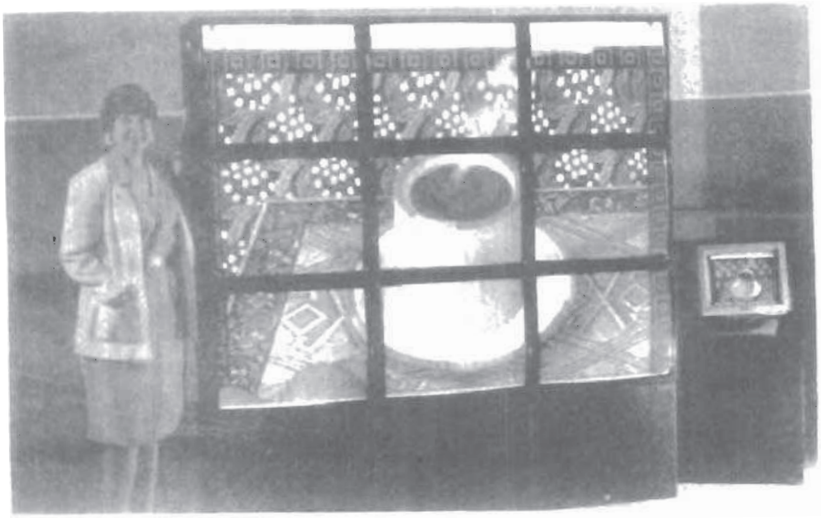


Figure 2-36
The MediaWall: A multiscreen display system. The image displayed on this 3-by-3 array of monitors was created by Deneba Software. (*Courtesy of RGB Spectrum.*)

Figure 2-37
Single- and dual-monitor graphics workstations. (*Courtesy of Intergraph Corporation.*)

Figures 2-38 and 2-39 illustrate examples of interactive graphics workstations containing multiple input and other devices. A typical setup for CAD applications is shown in Fig. 2-38. Various keyboards, button boxes, tablets, and mice are attached to the video monitors for use in the design process. Figure 2-39 shows features of some types of artist's workstations.

Figure 2-38
Multiple workstations for a CAD group. (*Courtesy of Hewlett-Packard Company.*)

Figure 2-39
An artist's workstation, featuring a color raster monitor, keyboard, graphics tablet with hand cursor, and a light table, in addition to data storage and telecommunications devices. (*Courtesy of DICOMED Corporation.*)

2-5
INPUT DEVICES

Various devices are available for data input on graphics workstations. Most systems have a keyboard and one or more additional devices specially designed for interactive input. These include a mouse, trackball, spaceball, joystick, digitizers,

dials, and button boxes. Some other input devices used in particular applications are data gloves, touch panels, image scanners, and voice systems.

Keyboards

An alphanumeric keyboard on a graphics system is used primarily as a device for entering text strings. The keyboard is an efficient device for inputting such nongraphic data as picture labels associated with a graphics display. Keyboards can also be provided with features to facilitate entry of screen coordinates, menu selections, or graphics functions.

Cursor-control keys and function keys are common features on general-purpose keyboards. Function keys allow users to enter frequently used operations in a single keystroke, and cursor-control keys can be used to select displayed objects or coordinate positions by positioning the screen cursor. Other types of cursor-positioning devices, such as a trackball or joystick, are included on some keyboards. Additionally, a numeric keypad is often included on the keyboard for fast entry of numeric data. Typical examples of general-purpose keyboards are given in Figs. 2-1, 2-33, and 2-34. Fig. 2-40 shows an ergonomic keyboard design.

For specialized applications, input to a graphics application may come from a set of buttons, dials, or switches that select data values or customized graphics operations. Figure 2-41 gives an example of a button box and a set of input dials. Buttons and switches are often used to input predefined functions, and dials are common devices for entering scalar values. Real numbers within some defined range are selected for input with dial rotations. Potentiometers are used to measure dial rotations, which are then converted to deflection voltages for cursor movement.

Mouse

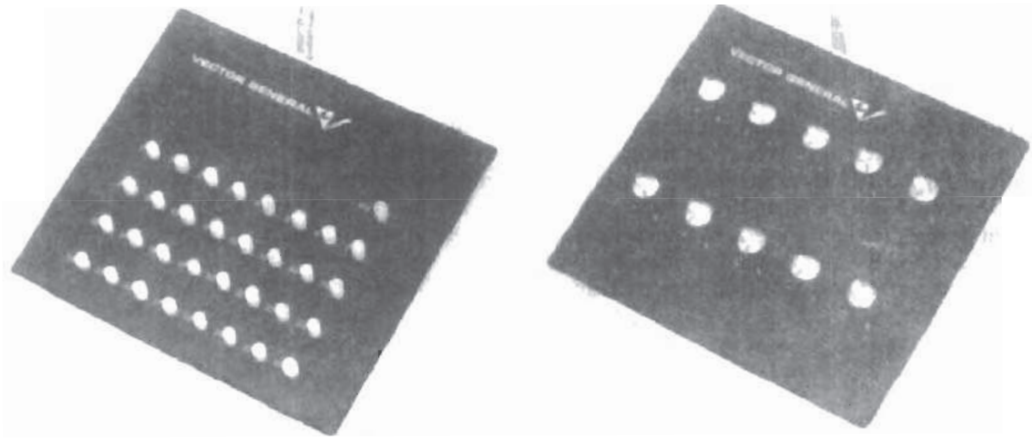
A **mouse** is small hand-held box used to position the screen cursor. Wheels or rollers on the bottom of the mouse can be used to record the amount and direc-

Figure 2-40
Ergonomically designed keyboard with removable palm rests. The slope of each half of the keyboard can be adjusted separately. *(Courtesy of Apple Computer, Inc.)*

tion of movement. Another method for detecting mouse motion is with an optical sensor. For these systems, the mouse is moved over a special mouse pad that has a grid of horizontal and vertical lines. The optical sensor detects movement across the lines in the grid.

Since a mouse can be picked up and put down at another position without change in cursor movement, it is used for making relative changes in the position of the screen cursor. One, two, or three buttons are usually included on the top of the mouse for signaling the execution of some operation, such as recording cursor position or invoking a function. Most general-purpose graphics systems now include a mouse and a keyboard as the major input devices, as in Figs. 2-1, 2-33, and 2-34.

Additional devices can be included in the basic mouse design to increase the number of allowable input parameters. The Z mouse in Fig. 2-42 includes



(a) (b)

Figure 2-41
A button box (a) and a set of input dials (b). *(Courtesy of Vector General.)*

Figure 2-42
The Z mouse features three buttons, a mouse ball underneath, a thumbwheel on the side, and a trackball on top. *(Courtesy of Multipoint Technology Corporation.)*

three buttons, a thumbwheel on the side, a trackball on the top, and a standard mouse ball underneath. This design provides six degrees of freedom to select spatial positions, rotations, and other parameters. With the Z mouse, we can pick up an object, rotate it, and move it in any direction, or we can navigate our viewing position and orientation through a three-dimensional scene. Applications of the Z mouse include virtual reality, CAD, and animation.

Trackball and Spaceball

As the name implies, a **trackball** is a ball that can be rotated with the fingers or palm of the hand, as in Fig. 2-43, to produce screen-cursor movement. Potentiometers, attached to the ball, measure the amount and direction of rotation. Trackballs are often mounted on keyboards (Fig. 2-15) or other devices such as the Z mouse (Fig. 2-42).

While a trackball is a two-dimensional positioning device, a **spaceball** (Fig. 2-45) provides six degrees of freedom. Unlike the trackball, a spaceball does not actually move. Strain gauges measure the amount of pressure applied to the spaceball to provide input for spatial positioning and orientation as the ball is pushed or pulled in various directions. Spaceballs are used for three-dimensional positioning and selection operations in virtual-reality systems, modeling, animation, CAD, and other applications.

Joysticks

A **joystick** consists of a small, vertical lever (called the stick) mounted on a base that is used to steer the screen cursor around. Most joysticks select screen positions with actual stick movement; others respond to pressure on the stick. Figure 2-44 shows a movable joystick. Some joysticks are mounted on a keyboard; others function as stand-alone units.

The distance that the stick is moved in any direction from its center position corresponds to screen-cursor movement in that direction. Potentiometers mounted at the base of the joystick measure the amount of movement, and springs return the stick to the center position when it is released. One or more buttons can be programmed to act as input switches to signal certain actions once a screen position has been selected.

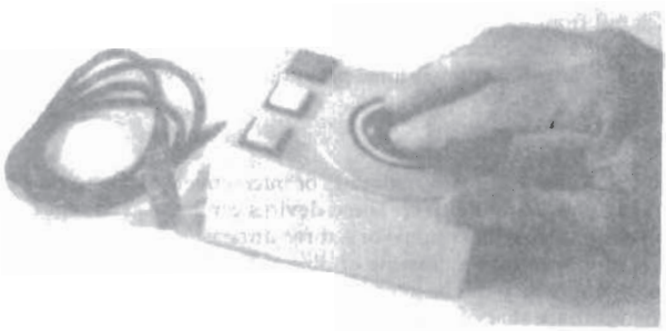


Figure 2-43
A three-button track ball. (*Courtesy of Measurement Systems Inc., Norwalk, Connecticut.*)

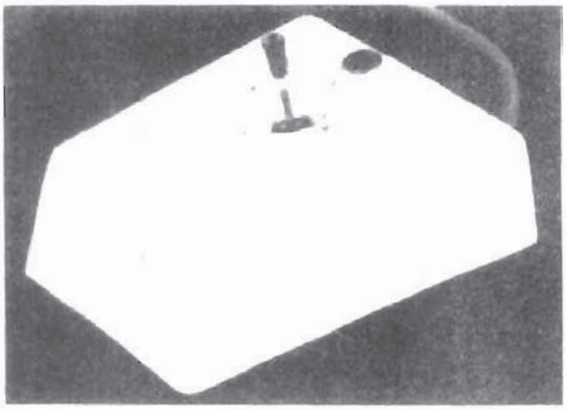


Figure 2-44
A moveable joystick. (*Courtesy of CalComp Group; Sanders Associates, Inc.*)

In another type of movable joystick, the stick is used to activate switches that cause the screen cursor to move at a constant rate in the direction selected. Eight switches, arranged in a circle, are sometimes provided, so that the stick can select any one of eight directions for cursor movement. Pressure-sensitive joysticks, also called isometric joysticks, have a nonmovable stick. Pressure on the stick is measured with strain gauges and converted to movement of the cursor in the direction specified.

Data Glove

Figure 2-45 shows a **data glove** that can be used to grasp a "virtual" object. The glove is constructed with a series of sensors that detect hand and finger motions. Electromagnetic coupling between transmitting antennas and receiving antennas is used to provide information about the position and orientation of the hand. The transmitting and receiving antennas can each be structured as a set of three mutually perpendicular coils, forming a three-dimensional Cartesian coordinate system. Input from the glove can be used to position or manipulate objects in a virtual scene. A two-dimensional projection of the scene can be viewed on a video monitor, or a three-dimensional projection can be viewed with a headset.

Digitizers

A common device for drawing, painting, or interactively selecting coordinate positions on an object is a **digitizer**. These devices can be used to input coordinate values in either a two-dimensional or a three-dimensional space. Typically, a digitizer is used to scan over a drawing or object and to input a set of discrete coordinate positions, which can be joined with straight-line segments to approximate the curve or surface shapes.

One type of digitizer is the **graphics tablet** (also referred to as a data tablet), which is used to input two-dimensional coordinates by activating a hand cursor or stylus at selected positions on a flat surface. A hand cursor contains cross hairs for sighting positions, while a stylus is a pencil-shaped device that is pointed at

Figure 2-45
A virtual-reality scene, displayed on a two-dimensional video monitor, with input from a data glove and a spaceball. *(Courtesy of The Computer Graphics Center, Darmstadt, Germany.)*

positions on the tablet. Figures 2-46 and 2-47 show examples of desktop and floor-model tablets, using hand cursors that are available with 2, 4, or 16 buttons. Examples of stylus input with a tablet are shown in Figs. 2-48 and 2-49. The artist's digitizing system in Fig. 2-49 uses electromagnetic resonance to detect the three-dimensional position of the stylus. This allows an artist to produce different brush strokes with different pressures on the tablet surface. Tablet size varies from 12 by 12 inches for desktop models to 44 by 60 inches or larger for floor models. Graphics tablets provide a highly accurate method for selecting coordinate positions, with an accuracy that varies from about 0.2 mm on desktop models to about 0.05 mm or less on larger models.

Many graphics tablets are constructed with a rectangular grid of wires embedded in the tablet surface. Electromagnetic pulses are generated in sequence

Figure 2-46
The SummaSketch III desktop tablet with a 16-button hand cursor. *(Courtesy of Summagraphics Corporation.)*

Figure 2-47
The Microgrid III tablet with a 16-button hand cursor, designed for digitizing larger drawings. (*Courtesy of Summagraphics Corporation.*)

Figure 2-48
The NotePad desktop tablet with stylus. (*Courtesy of CalComp Digitizer Division, a part of CalComp, Inc.*)

along the wires, and an electric signal is induced in a wire coil in an activated stylus or hand cursor to record a tablet position. Depending on the technology, either signal strength, coded pulses, or phase shifts can be used to determine the position on the tablet.

Acoustic (or sonic) tablets use sound waves to detect a stylus position. Either strip microphones or point microphones can be used to detect the sound emitted by an electrical spark from a stylus tip. The position of the stylus is calcu-

Figure 2-49
An artist's digitizer system, with a pressure-sensitive, cordless stylus. (*Courtesy of Wacom Technology Corporation.*)

lated by timing the arrival of the generated sound at the different microphone positions. An advantage of two-dimensional accoustic tablets is that the microphones can be placed on any surface to form the "tablet" work area. This can be convenient for various applications, such as digitizing drawings in a book.

Three-dimensional digitizers use sonic or electromagnetic transmissions to record positions. One electromagnetic transmission method is similar to that used in the data glove: A coupling between the transmitter and receiver is used to compute the location of a stylus as it moves over the surface of an object. Figure 2-50 shows a three-dimensional digitizer designed for Apple Macintosh computers. As the points are selected on a nonmetallic object, a wireframe outline of the surface is displayed on the computer screen. Once the surface outline is constructed, it can be shaded with lighting effects to produce a realistic display of the object. Resolution of this system is from 0.8 mm to 0.08 mm, depending on the model.

Image Scanners

Drawings, graphs, color and black-and-white photos, or text can be stored for computer processing with an **image scanner** by passing an optical scanning mechanism over the information to be stored. The gradations of gray scale or color are then recorded and stored in an array. Once we have the internal representation of a picture, we can apply transformations to rotate, scale, or crop the picture to a particular screen area. We can also apply various image-processing methods to modify the array representation of the picture. For scanned text input, various editing operations can be performed on the stored documents. Some scanners are able to scan either graphical representations or text, and they come in a variety of sizes and capabilities. A small hand-model scanner is shown in Fig. 2-51, while Figs 2-52 and 2-53 show larger models.

Figure 2-50
A three-dimensional digitizing system for use with Apple Macintosh computers. *(Courtesy of Mira Imaging.)*

Figure 2-51
A hand-held scanner that can be used to input either text or graphics images. (*Courtesy of Thunderware, Inc.*)

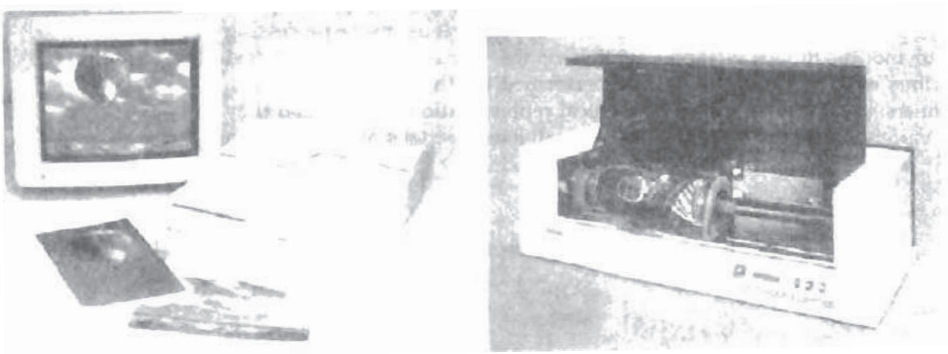


Figure 2-52
Desktop full-color scanners: (a) Flatbed scanner with a resolution of 600 dots per inch. (*Courtesy of Sharp Electronics Corporation.*) (b) Drum scanner with a selectable resolution from 50 to 4000 dots per inch. (*Courtesy of Howtek, Inc.*)

Touch Panels

As the name implies, **touch panels** allow displayed objects or screen positions to be selected with the touch of a finger. A typical application of touch panels is for the selection of processing options that are represented with graphical icons. Some systems, such as the plasma panels shown in Fig. 2-54, are designed with touch screens. Other systems can be adapted for touch input by fitting a transparent device with a touch-sensing mechanism over the video monitor screen. Touch input can be recorded using optical, electrical, or acoustical methods.

Optical touch panels employ a line of infrared light-emitting diodes (LEDs) along one vertical edge and along one horizontal edge of the frame. The opposite vertical and horizontal edges contain light detectors. These detectors are used to record which beams are interrupted when the panel is touched. The two crossing

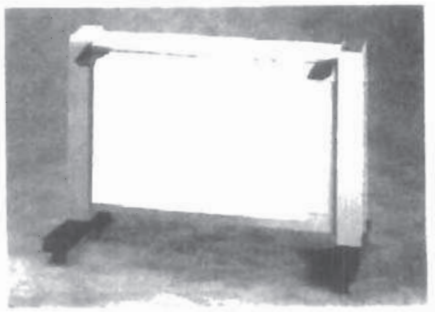


Figure 2-53
A large floor-model scanner used to scan architectural and engineering drawings up to 40 inches wide and 100 feet long. (*Courtesy of Summagraphics Corporation.*)

beams that are interrupted identify the horizontal and vertical coordinates of the screen position selected. Positions can be selected with an accuracy of about 1/4 inch. With closely spaced LEDs, it is possible to break two horizontal or two vertical beams simultaneously. In this case, an average position between the two interrupted beams is recorded. The LEDs operate at infrared frequencies, so that the light is not visible to a user. Figure 2-55 illustrates the arrangement of LEDs in an optical touch panel that is designed to match the color and contours of the system to which it is to be fitted.

An electrical touch panel is constructed with two transparent plates separated by a small distance. One of the plates is coated with a conducting material, and the other plate is coated with a resistive material. When the outer plate is touched, it is forced into contact with the inner plate. This contact creates a voltage drop across the resistive plate that is converted to the coordinate values of the selected screen position.

In acoustical touch panels, high-frequency sound waves are generated in the horizontal and vertical directions across a glass plate. Touching the screen causes part of each wave to be reflected from the finger to the emitters. The screen position at the point of contact is calculated from a measurement of the time interval between the transmission of each wave and its reflection to the emitter.

(a) (b)

Figure 2-54
Plasma panels with touch screens. (*Courtesy of Photonics Systems.*)

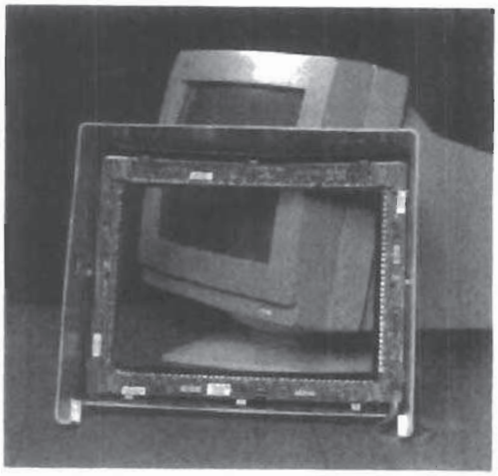


Figure 2-55
An optical touch panel, showing the arrangement of infrared LED units and detectors around the edges of the frame. *(Courtesy of Carroll Touch, Inc.)*

Light Pens

Figure 2-56 shows the design of one type of **light pen**. Such pencil-shaped devices are used to select screen positions by detecting the light coming from points on the CRT screen. They are sensitive to the short burst of light emitted from the phosphor coating at the instant the electron beam strikes a particular point. Other light sources, such as the background light in the room, are usually not detected by a light pen. An activated light pen, pointed at a spot on the screen as the electron beam lights up that spot, generates an electrical pulse that causes the coordinate position of the electron beam to be recorded. As with cursor-positioning devices, recorded light-pen coordinates can be used to position an object or to select a processing option.

Although light pens are still with us, they are not as popular as they once were since they have several disadvantages compared to other input devices that have been developed. For one, when a light pen is pointed at the screen, part of the screen image is obscured by the hand and pen. And prolonged use of the light pen can cause arm fatigue. Also, light pens require special implementations for some applications because they cannot detect positions within black areas. To be able to select positions in any screen area with a light pen, we must have some nonzero intensity assigned to each screen pixel. In addition, light pens sometimes give false readings due to background lighting in a room.

Voice Systems

Speech recognizers are used in some graphics workstations as input devices to accept voice commands The **voice-system** input can be used to initiate graphics

Figure 2-56
A light pen activated with a button switch. (*Courtesy of Interactive Computer Products.*)

operations or to enter data. These systems operate by matching an input against a predefined dictionary of words and phrases.

A dictionary is set up for a particular operator by having the operator speak the command words to be used into the system. Each word is spoken several times, and the system analyzes the word and establishes a frequency pattern for that word in the dictionary along with the corresponding function to be performed. Later, when a voice command is given, the system searches the dictionary for a frequency-pattern match. Voice input is typically spoken into a microphone mounted on a headset, as in Fig. 2-57. The microphone is designed to minimize input of other background sounds. If a different operator is to use the system, the dictionary must be reestablished with that operator's voice patterns. Voice systems have some advantage over other input devices, since the attention of the operator does not have to be switched from one device to another to enter a command.

Figure 2-57
A speech-recognition system. (*Courtesy of Threshold Technology, Inc.*)

2-6
HARD-COPY DEVICES

We can obtain hard-copy output for our images in several formats. For presentations or archiving, we can send image files to devices or service bureaus that will produce 35-mm slides or overhead transparencies. To put images on film, we can simply photograph a scene displayed on a video monitor. And we can put our pictures on paper by directing graphics output to a printer or plotter.

The quality of the pictures obtained from a device depends on dot size and the number of dots per inch, or lines per inch, that can be displayed. To produce smooth characters in printed text strings, higher-quality printers shift dot positions so that adjacent dots overlap.

Printers produce output by either impact or nonimpact methods. *Impact* printers press formed character faces against an inked ribbon onto the paper. A line printer is an example of an impact device, with the typefaces mounted on bands, chains, drums, or wheels. *Nonimpact* printers and plotters use laser techniques, ink-jet sprays, xerographic processes (as used in photocopying machines), electrostatic methods, and electrothermal methods to get images onto paper.

Character impact printers often have a *dot-matrix* print head containing a rectangular array of protruding wire pins, with the number of pins depending on the quality of the printer. Individual characters or graphics patterns are obtained by retracting certain pins so that the remaining pins form the pattern to be printed. Figure 2-58 shows a picture printed on a dot-matrix printer.

In a *laser* device, a laser beam creates a charge distribution on a rotating drum coated with a photoelectric material, such as selenium. Toner is applied to the drum and then transferred to paper. Figure 2-59 shows examples of desktop laser printers with a resolution of 360 dots per inch.

Ink-jet methods produce output by squirting ink in horizontal rows across a roll of paper wrapped on a drum. The electrically charged ink stream is deflected by an electric field to produce dot-matrix patterns. A desktop ink-jet plotter with

Figure 2-58
A picture generated on a dot-matrix printer showing how the density of the dot patterns can be varied to produce light and dark areas. (*Courtesy of Apple Computer, Inc.*)

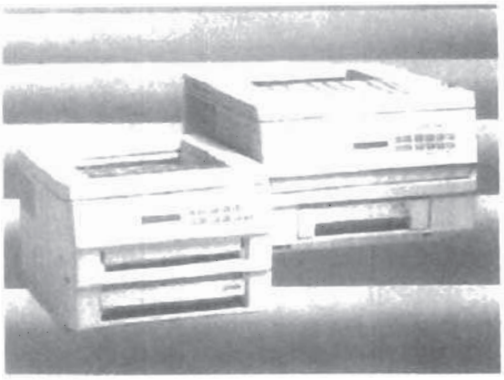


Figure 2-59
Small-footprint laser printers.
(Courtesy of Texas Instruments.)

a resolution of 360 dots per inch is shown in Fig. 2-60, and examples of larger high-resolution ink-jet printer/plotters are shown in Fig. 2-61.

An *electrostatic* device places a negative charge on the paper, one complete row at a time along the length of the paper. Then the paper is exposed to a toner. The toner is positively charged and so is attracted to the negatively charged areas, where it adheres to produce the specified output. A color electrostatic printer/plotter is shown in Fig. 2-62. *Electrothermal* methods use heat in a dot-matrix print head to output patterns on heat-sensitive paper.

We can get limited color output on an impact printer by using different-colored ribbons. Nonimpact devices use various techniques to combine three color pigments (cyan, magenta, and yellow) to produce a range of color patterns. Laser and xerographic devices deposit the three pigments on separate passes; ink-jet methods shoot the three colors simultaneously on a single pass along each print line on the paper.

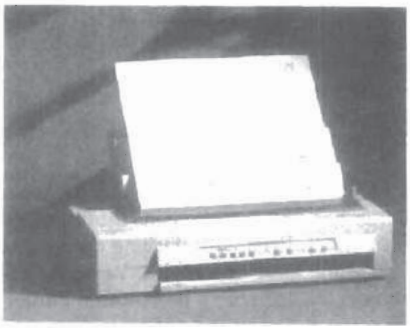
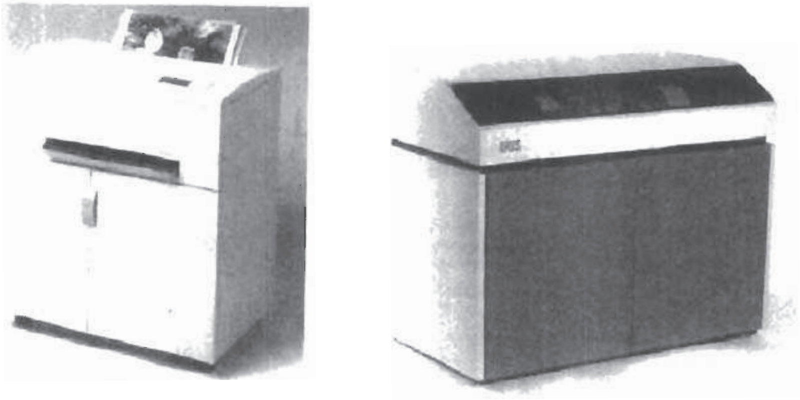


Figure 2-60
A 360-dot-per-inch desktop ink-jet plotter. *(Courtesy of Summagraphics Corporation.)*



(a) (b)

Figure 2-61
Floor-model, ink-jet color printers that use variable dot size to achieve an equivalent resolution of 1500 to 1800 dots per inch. (*Courtesy of IRIS Graphics Inc., Bedford, Massachusetts.*)

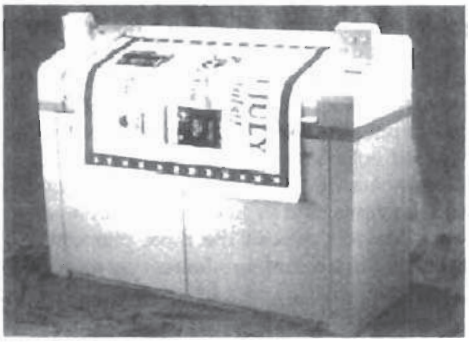


Figure 2-62
An electrostatic printer that can display 400 dots per inch. (*Courtesy of CalComp Digitizer Division, a part of CalComp, Inc.*)

Drafting layouts and other drawings are typically generated with ink-jet or pen plotters. A pen plotter has one or more pens mounted on a carriage, or crossbar, that spans a sheet of paper. Pens with varying colors and widths are used to produce a variety of shadings and line styles. Wet-ink, ball-point, and felt-tip pens are all possible choices for use with a pen plotter. Plotter paper can lie flat or be rolled onto a drum or belt. Crossbars can be either moveable or stationary, while the pen moves back and forth along the bar. Either clamps, a vacuum, or an electrostatic charge hold the paper in position. An example of a table-top flatbed pen plotter is given in Figure 2-63, and a larger, rollfeed pen plotter is shown in Fig. 2-64.

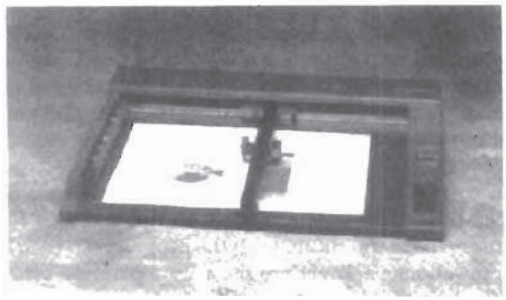


Figure 2-63
A desktop pen plotter with a resolution of 0.025 mm. (*Courtesy of Summagraphics Corporation.*)

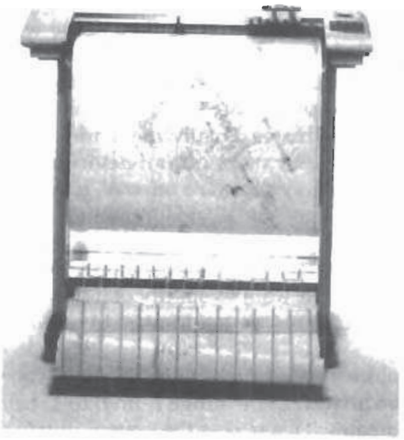


Figure 2-64
A large, rollfeed pen plotter with automatic multicolor 8-pen changer and a resolution of 0.0127 mm. (*Courtesy of Summagraphics Corporation.*)

2-7 GRAPHICS SOFTWARE

There are two general classifications for graphics software: general programming packages and special-purpose applications packages. A general graphics programming package provides an extensive set of graphics functions that can be

used in a high-level programming language, such as C or FORTRAN. An example of a general graphics programming package is the GL (Graphics Library) system on Silicon Graphics equipment. Basic functions in a general package include those for generating picture components (straight lines, polygons, circles, and other figures), setting color and intensity values, selecting views, and applying transformations. By contrast, application graphics packages are designed for nonprogrammers, so that users can generate displays without worrying about how graphics operations work. The interface to the graphics routines in such packages allows users to communicate with the programs in their own terms. Examples of such applications packages are the artist's painting programs and various business, medical, and CAD systems.

Coordinate Representations

With few exceptions, general graphics packages are designed to be used with Cartesian coordinate specifications. If coordinate values for a picture are specified in some other reference frame (spherical, hyberbolic, etc.), they must be converted to Cartesian coordinates before they can be input to the graphics package. Special-purpose packages may allow use of other coordinate frames that are appropriate to the application. In general, several different Cartesian reference frames are used to construct and display a scene. We can construct the shape of individual objects, such as trees or furniture, in a scene within separate coordinate reference frames called **modeling coordinates**, or sometimes **local coordinates** or **master coordinates**. Once individual object shapes have been specified, we can place the objects into appropriate positions within the scene using a reference frame called **world coordinates**. Finally, the world-coordinate description of the scene is transferred to one or more output-device reference frames for display. These display coordinate systems are referred to as **device coordinates**, or **screen coordinates** in the case of a video monitor. Modeling and world-coordinate definitions allow us to set any convenient floating-point or integer dimensions without being hampered by the constraints of a particular output device. For some scenes, we might want to specify object dimensions in fractions of a foot, while for other applications we might want to use millimeters, kilometers, or light-years.

Generally, a graphics system first converts world-coordinate positions to **normalized device coordinates**, in the range from 0 to 1, before final conversion to specific device coordinates. This makes the system independent of the various devices that might be used at a particular workstation. Figure 2-65 illustrates the sequence of coordinate transformations from modeling coordinates to device coordinates for a two-dimensional application. An initial modeling-coordinate position (x_{mc}, y_{mc}) in this illustration is transferred to a device coordinate position (x_{dc}, y_{dc}) with the sequence:

$$(x_{mc}, y_{mc}) \rightarrow (x_{wc}, y_{wc}) \rightarrow (x_{nc}, y_{nc}) \rightarrow (x_{dc}, y_{dc})$$

The modeling and world-coordinate positions in this transformation can be any floating-point values; normalized coordinates satisfy the inequalities: $0 \leq x_{nc} \leq 1$, $0 \leq y_{nc} \leq 1$; and the device coordinates x_{dc} and y_{dc} are integers within the range (0, 0) to (x_{max}, y_{max}) for a particular output device. To accommodate differences in scales and aspect ratios, normalized coordinates are mapped into a square area of the output device so that proper proportions are maintained.

Graphics Functions

A general-purpose graphics package provides users with a variety of functions for creating and manipulating pictures. These routines can be categorized according to whether they deal with output, input, attributes, transformations, viewing, or general control.

The basic building blocks for pictures are referred to as **output primitives**. They include character strings and geometric entities, such as points, straight lines, curved lines, filled areas (polygons, circles, etc.), and shapes defined with arrays of color points. Routines for generating output primitives provide the basic tools for constructing pictures.

Attributes are the properties of the output primitives; that is, an attribute describes how a particular primitive is to be displayed. They include intensity and color specifications, line styles, text styles, and area-filling patterns. Functions within this category can be used to set attributes for an individual primitive class or for groups of output primitives.

We can change the size, position, or orientation of an object within a scene using **geometric transformations**. Similar **modeling transformations** are used to construct a scene using object descriptions given in modeling coordinates.

Given the primitive and attribute definition of a picture in world coordinates, a graphics package projects a selected view of the picture on an output device. **Viewing transformations** are used to specify the view that is to be presented and the portion of the output display area that is to be used.

Pictures can be subdivided into component parts, called **structures** or **segments** or **objects**, depending on the software package in use. Each structure defines one logical unit of the picture. A scene with several objects could reference each individual object in a separate named structure. Routines for processing

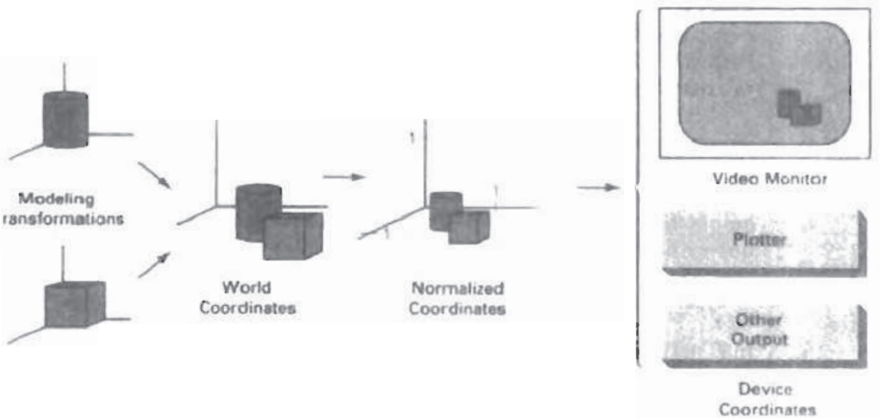


Figure 2-65
The transformation sequence from modeling coordinates to device coordinates for a two-dimensional scene. Object shapes are defined in local modeling-coordinate systems, then positioned within the overall world-coordinate scene. World-coordinate specifications are then transformed into normalized coordinates. At the final step, individual device drivers transfer the normalized-coordinate representation of the scene to the output devices for display.

structures carry out operations such as the creation, modification, and transformation of structures.

Interactive graphics applications use various kinds of input devices, such as a mouse, a tablet, or a joystick. **Input functions** are used to control and process the data flow from these interactive devices.

Finally, a graphics package contains a number of housekeeping tasks, such as clearing a display screen and initializing parameters. We can lump the functions for carrying out these chores under the heading **control operations**.

Software Standards

The primary goal of standardized graphics software is portability. When packages are designed with standard graphics functions, software can be moved easily from one hardware system to another and used in different implementations and applications. Without standards, programs designed for one hardware system often cannot be transferred to another system without extensive rewriting of the programs.

International and national standards planning organizations in many countries have cooperated in an effort to develop a generally accepted standard for computer graphics. After considerable effort, this work on standards led to the development of the **Graphical Kernel System (GKS)**. This system was adopted as the first graphics software standard by the International Standards Organization (ISO) and by various national standards organizations, including the American National Standards Institute (ANSI). Although GKS was originally designed as a two-dimensional graphics package, a three-dimensional GKS extension was subsequently developed. The second software standard to be developed and approved by the standards orgainzations was **PHIGS (Programmer's Hierarchical Interactive Graphics Standard)**, which is an extension of GKS. Increased capabilities for object modeling, color specifications, surface rendering, and picture manipulations are provided in PHIGS. Subsequently, an extension of PHIGS, called PHIGS+, was developed to provide three-dimensional surface-shading capabilities not available in PHIGS.

Standard graphics functions are defined as a set of specifications that is independent of any programming language. A **language binding** is then defined for a particular high-level programming language. This binding gives the syntax for accessing the various standard graphics functions from this language. For example, the general form of the PHIGS (and GKS) function for specifying a sequence of $n - 1$ connected two-dimensional straight line segments is

```
polyline(n, x, y)
```

In FORTRAN, this procedure is implemented as a subroutine with the name *GPL*. A graphics programmer, using FORTRAN, would invoke this procedure with the subroutine call statement `CALL GPL(N, X, Y)`, where `X` and `Y` are one-dimensional arrays of coordinate values for the line endpoints. In C, the procedure would be invoked with `ppolyline(n, pts)`, where `pts` is the list of coordinate endpoint positions. Each language binding is defined to make best use of the corresponding language capabilities and to handle various syntax issues, such as data types, parameter passing, and errors.

In the following chapters, we use the standard functions defined in PHIGS as a framework for discussing basic graphics concepts and the design and application of graphics packages. Example programs are presented in Pascal to illus-

trate the algorithms for implementation of the graphics functions and to illustrate also some applications of the functions. Descriptive names for functions, based on the PHIGS definitions, are used whenever a graphics function is referenced in a program.

Although PHIGS presents a specification for basic graphics functions, it does not provide a standard methodology for a graphics interface to output devices. Nor does it specify methods for storing and transmitting pictures. Separate standards have been developed for these areas. Standardization for device interface methods is given in the **Computer Graphics Interface (CGI)** system. And the **Computer Graphics Metafile (CGM)** system specifies standards for archiving and transporting pictures.

PHIGS Workstations

Generally, the term *workstation* refers to a computer system with a combination of input and output devices that is designed for a single user. In PHIGS and GKS, however, the term **workstation** is used to identify various combinations of graphics hardware and software. A PHIGS workstation can be a single output device, a single input device, a combination of input and output devices, a file, or even a window displayed on a video monitor.

To define and use various "workstations" within an applications program, we need to specify a *workstation identifier* and the workstation type. The following statements give the general structure of a PHIGS program:

```
openPhigs (errorFile, memorySize)
openWorkstation (ws, connection, type)
    { create and display picture}
closeWorkstation (ws)
closePhigs
```

where parameter `errorFile` is to contain any error messages that are generated, and parameter `memorySize` specifies the size of an internal storage area. The workstation identifier (an integer) is given in parameter `ws`, and parameter `connection` states the access mechanism for the workstation. Parameter `type` specifies the particular category for the workstation, such as an *input* device, an *output* device, a combination *outin* device, or an input or output metafile.

Any number of workstations can be open in a particular application, with input coming from the various open input devices and output directed to all the open output devices. We discuss input and output methods in applications programs in Chapter 8, after we have explored the basic procedures for creating and manipulating pictures.

SUMMARY

In this chapter, we have surveyed the major hardware and software features of computer graphics systems. Hardware components include video monitors, hard-copy devices, keyboards, and other devices for graphics input or output. Graphics software includes special applications packages and general programming packages.

The predominant graphics display device is the raster refresh monitor, based on television technology. A raster system uses a frame buffer to store intensity information for each screen position (pixel). Pictures are then painted on the

screen by retrieving this information from the frame buffer as the electron beam in the CRT sweeps across each scan line, from top to bottom. Older vector displays construct pictures by drawing lines between specified line endpoints. Picture information is then stored as a set of line-drawing instructions.

Many other video display devices are available. In particular, flat-panel display technology is developing at a rapid rate, and these devices may largely replace raster displays in the near future. At present, flat-panel displays are commonly used in small systems and in special-purpose systems. Flat-panel displays include plasma panels and liquid-crystal devices. Although vector monitors can be used to display high-quality line drawings, improvements in raster display technology have caused vector monitors to be largely replaced with raster systems.

Other display technologies include three-dimensional and stereoscopic viewing systems. Virtual-reality systems can include either a stereoscopic headset or a standard video monitor.

For graphical input, we have a range of devices to choose from. Keyboards, button boxes, and dials are used to input text, data values, or programming options. The most popular "pointing" device is the mouse, but trackballs, spaceballs, joysticks, cursor-control keys, and thumbwheels are also used to position the screen cursor. In virtual-reality environments, data gloves are commonly used. Other input devices include image scanners, digitizers, touch panels, light pens, and voice systems.

Hard-copy devices for graphics workstations include standard printers and plotters, in addition to devices for producing slides, transparencies, and film output. Printing methods include dot matrix, laser, ink jet, electrostatic, and electrothermal. Plotter methods include pen plotting and combination printer-plotter devices.

Graphics software can be roughly classified as applications packages or programming packages. Applications graphics software include CAD packages, drawing and painting programs, graphing packages, and visualization programs. Common graphics programming packages include PHIGS, PHIGS+, GKS, 3D GKS, and GL. Software standards, such as PHIGS, GKS, CGI, and CGM, are evolving and are becoming widely available on a variety of machines.

Normally, graphics packages require coordinate specifications to be given with respect to Cartesian reference frames. Each object for a scene can be defined in a separate modeling Cartesian coordinate system, which is then mapped to world coordinates to construct the scene. From world coordinates, objects are transferred to normalized device coordinates, then to the final display device coordinates. The transformations from modeling coordinates to normalized device coordinates are independent of particular devices that might be used in an application. Device drivers are then used to convert normalized coordinates to integer device coordinates.

Functions in graphics programming packages can be divided into the following categories: output primitives, attributes, geometric and modeling transformations, viewing transformations, structure operations, input functions, and control operations.

Some graphics systems, such as PHIGS and GKS, use the concept of a "workstation" to specify devices or software that are to be used for input or output in a particular application. A workstation identifier in these systems can refer to a file; a single device, such as a raster monitor; or a combination of devices, such as a monitor, keyboard, and a mouse. Multiple workstations can be open to provide input or to receive output in a graphics application.

REFERENCES

A general treatment of electronic displays, including flat-panel devices, is available in Sherr (1993). Flat-panel devices are discussed in Depp and Howard (1993). Tannas (1985) provides a reference for both flat-panel displays and CRTs. Additional information on raster-graphics architecture can be found in Foley, et al. (1990). Three-dimensional terminals are discussed in Fuchs et al. (1982), Johnson (1982), and Ikedo (1984). Head-mounted displays and virtual-reality environments are discussed in Chung et al. (1989).

For information on PHIGS and PHIGS+, see Hopgood and Duce (1991), Howard et al. (1991), Gaskins (1992), and Blake (1993). Information on the two-dimensional GKS standard and on the evolution of graphics standards is available in Hopgood et al. (1983). An additional reference for GKS is Enderle, Kansy, and Pfaff (1984).

EXERCISES

2-1. List the operating characteristics for the following display technologies: raster refresh systems, vector refresh systems, plasma panels, and LCDs.

2-2. List some applications appropriate for each of the display technologies in Exercise 2-1.

2-3. Determine the resolution (pixels per centimeter) in the x and y directions for the video monitor in use on your system. Determine the aspect ratio, and explain how relative proportions of objects can be maintained on your system.

2-4. Consider three different raster systems with resolutions of 640 by 480, 1280 by 1024, and 2560 by 2048. What size frame buffer (in bytes) is needed for each of these systems to store 12 bits per pixel? How much storage is required for each system if 24 bits per pixel are to be stored?

2-5. Suppose an RGB raster system is to be designed using an 8-inch by 10-inch screen with a resolution of 100 pixels per inch in each direction. If we want to store 6 bits per pixel in the frame buffer, how much storage (in bytes) do we need for the frame buffer?

2-6. How long would it take to load a 640 by 480 frame buffer with 12 bits per pixel, if 10^5 bits can be transferred per second? How long would it take to load a 24-bit per pixel frame buffer with a resolution of 1280 by 1024 using this same transfer rate?

2-7. Suppose we have a computer with 32 bits per word and a transfer rate of 1 mip (one million instructions per second). How long would it take to fill the frame buffer of a 300-dpi (dot per inch) laser printer with a page size of 8 1/2 inches by 11 inches?

2-8. Consider two raster systems with resolutions of 640 by 480 and 1280 by 1024. How many pixels could be accessed per second in each of these systems by a display controller that refreshes the screen at a rate of 60 frames per second? What is the access time per pixel in each system?

2-9. Suppose we have a video monitor with a display area that measures 12 inches across and 9.6 inches high. If the resolution is 1280 by 1024 and the aspect ratio is 1, what is the diameter of each screen point?

2-10. How much time is spent scanning across each row of pixels during screen refresh on a raster system with a resolution of 1280 by 1024 and a refresh rate of 60 frames per second?

2-11. Consider a noninterlaced raster monitor with a resolution of n by m (m scan lines and n pixels per scan line), a refresh rate of r frames per second, a horizontal retrace time of t_{horiz}, and a vertical retrace time of t_{vert}. What is the fraction of the total refresh time per frame spent in retrace of the electron beam?

2-12. What is the fraction of the total refresh time per frame spent in retrace of the electron beam for a noninterlaced raster system with a resolution of 1280 by 1024, a refresh rate of 60 Hz, a horizontal retrace time of 5 microseconds, and a vertical retrace time of 500 microseconds?

2-13. Assuming that a certain full-color (24-bit per pixel) RGB raster system has a 512-by-512 frame buffer, how many distinct color choices (intensity levels) would we have available? How many different colors could we display at any one time?

2-14. Compare the advantages and disadvantages of a three-dimensional monitor using a varifocal mirror with a stereoscopic system.

2-15. List the different input and output components that are typically used with virtual-reality systems. Also explain how users interact with a virtual scene displayed with different output devices, such as two-dimensional and stereoscopic monitors.

2-16. Explain how virtual-reality systems can be used in design applications. What are some other applications for virtual-reality systems?

2-17. List some applications for large-screen displays.

2-18. Explain the differences between a general graphics system designed for a programmer and one designed for a specific application, such as architectural design?